AD-A120 838 ON FORECASTING WITH UNIVARIATE AUTOREGRESSIVE PROCESSES: A BAYESIAN APPRO..(U) OKLAHOMA STATE UNIV STILLWATER DEPT OF STATISTICS L BROEMELING ET AL. 1/1

UNCLASSIFIED 26 JUL 82 OSU-TR-37 N00014-82-K-0292 F/G 12/1 NL

END

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Oklahoma State University

Stillwater

(12)

TECHNICAL REPORT NO. 37
Department of Statistics
Oklahoma State University
Stillwater, Oklahoma 74078
July 26, 1982

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: OSU Tech Report #37	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): On Forecasting with univariate auto-regressive processes: A Bayesian approach.	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Lyle Broemeling, Margaret Land	8. CONTRACT OR GRANT NUMBER(s): N00014-82-K-0292
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Statistics Department, Oklahoma State University, Stillwater, Oklahoma 74078	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Department of the Navy, Office of Naval Research, Arlington, Virginia 22217	12. REPORT DATE: JULY 1982
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bayesian Inference; predictive distribution; autoregressive process; future observations, forecasting; normal-gamma prior density.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Using a normal-gamma prior density for the parameters of a p-th order autoregressive process, the Bayesian predictive density of k future observations is derived. It is shown that the joint predictive density of k future observations is the product of k univariate t densities. Our results are illustrated with one step-ahead forecasts of an AR(1) model.

ON FORECASTING WITH UNIVARIATE AUTOREGRESSIVE PROCESSES: A BAYESIAN APPROACH

LYLE BROEMELING
Oklahoma State University, Stillwater, Oklahoma 74074, U.S.A.

MARGARET LAND
Texas A and I University, Kingsville, Texas, 78363, U.S.A.

Using a normal-gamma prior density for the parameters of a p-th order autoregressive process, the Bayesian predictive density of k future observations is derived. It is shown that the joint predictive density of k future observations may be expressed as the product of k univariate t densities. Our results are illustrated with one-step ahead forecasts employing an AR(1) model with a conjugate prior density for the parameters.

-A-

1. Introduction

Consider a p-th order autoregressive model

$$Y(t) = \sum_{j=1}^{p} \theta_j Y(t-j) + \varepsilon(t) , \tag{1}$$

where $t = 1,2,\ldots,n$, $\theta_j \varepsilon R$, $Y(t)$ is the observation at time t, and $\varepsilon(1), \varepsilon(2),\ldots,\varepsilon(n)$ are n.i.d. $(0,\tau)$, $\tau > 0$. The parameters $\theta_1,\theta_2,\ldots,\theta_p$, and τ are unknown and the 'initial' observations $Y(0),Y(-1),\ldots,y(1-p)$, are assumed to be known constants.

Given a sample $S_n = [Y(1),\ldots,Y(n)]'$ of observations, how does one forecast future observations $W(j) = Y(n+j)$, where $j = 1,2,\ldots,k$, and k is an integer such that $k \geq 1$? Of course there are many non-Bayesian techniques of forecasting in the literature including Box and Jenkins, exponential smoothing, and stepwise autoregression, all of which are explained in Granger and Newbold (1977). In addition, Bayesian techniques of forecasting have been developed by Harrison and Stevens (1976), Zellner (1971) and Chow (1975), who base their forecasts on the Bayesian predictive distribution of the future observations.

The Bayesian predictive distribution is the conditional distribution of the future observations $W_k = [W(1),\ldots,W(k)]'$ given the past observations S_n and plays a prominent role in Bayesian methodology. Aitchison and Dunsmore (1975) develop the Bayesian predictive density for many of the traditional statistical models, but curiously, it has not often been used in time series analysis, except by Zellner and Chow, where the former author uses it with first and second order autoregressive processes

for $K = 1$ (one-step ahead prediction), and Chow derives the predictive moments for the general case.

The purpose of this study is to characterize the predictive distribution of W_k given S_n. It will be shown that when $k = 1$. The predictive distribution of W_1 is a univariate t and that when $k = 2$, the conditional predictive distribution of $W(2)$ given $W(1)$ is a univariate t and that the marginal predictive distribution of $W(1)$ is also a t . In general, assuming a normal-gamma conjugate prior density for the parameters, the predictive density of W_k is a product of univariate t densities.

This study is concluded with a numerial demonstration of one-step ahead forecasting with a first order autoregressive model. Using a normal-gamma prior density for the two parameters of the model, the mean and variance of the predictive distribution of $W(1) = Y(n + 1)$ is computed for a wide variety of models, sample sizes, and values of the prior mean of the autoregressive parameter.

2. The Prior and Posterior Analyses

Using the Bayesian approach, one must specify a prior density for $\theta = (\theta_1,\theta_2,\ldots,\theta_p)'$ and τ, and it often is convenient to use either a Jeffreys' improper prior

$$\xi_1(\theta,\tau) \propto 1/\tau,\ \tau > 0,\ \theta\epsilon R^p \qquad (2)$$

or the normal-gamma conjugate prior density

$$\xi_2(\theta,\tau) = \xi_{21}(\theta/\tau)\xi_{22}(\tau),\ \tau > 0,\ \theta\epsilon R^p\ , \qquad (3)$$

where the conditional prior density of θ given τ is

$$\xi_{21}(\theta/\tau) \propto \tau^{p/2} e^{-\frac{\tau}{2}(\theta-\mu)' \mathbb{P}(\theta-\mu)}, \quad \theta \varepsilon R^p \tag{4}$$

and

$$\xi_{22}(\tau) \propto \tau^{\alpha-1} e^{-\tau\beta}, \quad \tau > 0 \tag{5}$$

is the marginal prior density of τ.

Thus, apriori, θ given τ has a normal distribution with mean vector μ and precision matrix $\tau\mathbb{P}$, where $\mathbb{P}$ is a symmetric positive definite matrix, and τ has a gamma distribution with parameters $\alpha > 0$ and $\beta > 0$. This implies the marginal prior density of θ is

$$\xi_3(\theta) \propto [2\beta + (\theta-\mu)' \mathbb{P}(\theta-\mu)]^{-(1+2\alpha)/2}, \quad \theta \varepsilon R^p \tag{6}$$

which is a t density with 2α degrees of freedom, location μ, and precision matrix $(2\alpha)\mathbb{P}(2\beta)^{-1}$. Note, the parameters of the marginal prior distribution of τ are also parameters of the marginal prior density of θ, hence one's prior information about θ depends on one's prior opinion of τ.

The prior density ξ_2 of the parameters θ and τ is combined with the conditional density of the observations S_n given θ and τ, which is

$$f(S_n/\theta,\tau) \propto \tau^{n/2} \exp - \frac{\tau}{2} \sum_{t=1}^{n} [Y(t) - \sum_{j=1}^{p} \theta_j Y(t-j)]^2, S_n \varepsilon R^n \tag{7}$$

the product is the posterior density of θ and τ, namely,

$$\xi(\theta,\tau/S_n) \propto \tau^{\frac{n+p+2\alpha}{2}-1} \exp - \frac{\tau}{2} \{2\beta + (\theta-\mu)'\mathbb{P}(\theta-\mu) +$$

$$\sum_{t=1}^{n} [Y(t) - \sum_{j=1}^{p} \theta_j Y(t-j)]^2\}, \tag{8}$$

3. The Bayesian Predictive Distribution

The Bayesian predictive density of W_k (conditional on S_n) is

$$g(W_k/S_n) = \int_\Omega \xi(\theta,\tau/S_n)f(W_k/S_n,\theta,\tau)d\theta d\tau,\ W_k \varepsilon R^k \quad (9)$$

where $\Omega = \{(\theta,\tau):\theta\varepsilon R^p, \tau > 0\}$ and $f(W_k/S_n,\theta,\tau)$ is the conditional density of the k future observations W_k given S_n,θ, and τ. It is seen that the Bayesian predictive density of W_k is the average (with respect to the posterior distribution of the parameters) of the conditional predictive density of W_k given θ and τ.

The integrand of (9) is proportional to $\xi_2(\theta,\tau)f(S_n/\theta,\tau)f(W_k/S_n,\theta,\tau)$, thus the predictive density of W_k is the average (with respect to the prior distribution) of the distribution of S_n and W_k given θ and τ, which has density,

$$f(S_n,W_k/\theta,\tau) \propto \tau^{\frac{n+k}{2}} \exp -\frac{\tau}{2}\left\{\sum_{t=1}^{n}[Y(t)-\sum_{j=1}^{p}\theta_j Y(t-j)]^2 + \sum_{s=1}^{k}[W(s)-\sum_{j=1}^{p}\theta_j W(s-j)]^2\right\} \quad (10)$$

where $S_n \varepsilon R^n$, $W_k \varepsilon R^k$, and $W(-i) = Y(n-i)$, $i = 0,1,2,\ldots,$.

The joint distribution of S_n, W_k, θ, and τ is

$$f(S_n,W_k,\theta,\tau) = f(S_n,W_k/\theta,\tau)\xi_2(\theta,\tau), S_n\varepsilon R^n, W_k\varepsilon R^k, \theta\varepsilon R^p, \tau > 0, \quad (11)$$

thus

$$f(S_n,W_k) = \int_\Omega f(S_n,W_k,\theta,\tau)d\theta d\tau, S_n\varepsilon R^n, W_k\varepsilon R^k \quad (12)$$

is the joint density of the past and future observations, and the Bayesian predictive distribution of W_k will be identified from this density, because

$$g(W_k/S_n) \propto f(S_n,W_k),\ W_k\varepsilon R^k \quad (13)$$

It can be shown that if $K \geq 1$,

$$g(W_k/S_n) = g_1(W_{k-1},S_n)g_2(W_k,S_n), W_k \varepsilon R^k, \tag{14}$$

where

$$g_1(W_{k-1},S_n) \propto |A|^{-1/2}, W_{k-1} \varepsilon R^{k-1}, \tag{15}$$

$$g_2(W_k,S_n) \propto [C-B'A^{-1}B]^{-(n+2\alpha+k)/2}, W_k \varepsilon R^k, \tag{16}$$

and W_o does not depend on $W(1), W(2)$, etc.

The other quantities are

$$A = A_1 + A_2 + \mathbb{P}, \tag{17}$$

$$B = B_1 + B_2 + \mathbb{P}\mu, \tag{18}$$

and

$$C = \sum_{t=1}^{n} Y^2(t) + \sum_{s=1}^{k} W^2(s) + 2\beta + \mu' \mathbb{P}\mu, \tag{19}$$

where A_1 and A_2 are the pxp matrices

$$A_1 = [\sum_{t=1}^{n} Y(t-j)Y(t-\ell)], \ 1 \leq j \leq \ell \leq p \tag{20}$$

and

$$A_2 = [\sum_{s=1}^{k} W(s-j)W(s-\ell)], \ 1 \leq j \leq \ell \leq p \ . \tag{21}$$

The px1 vectors B_1 and B_2 are given by

$$B_1 = [\sum_{t=1}^{n} Y(t)Y(t-j)], \ 1 \leq j \leq p \tag{22}$$

and

$$B_2 = [\sum_{s=1}^{k} W(s)W(s-j), \ 1 \leq j \leq p \ . \tag{23}$$

Consider equations (14),(15), and (16), then if $k = 1$, g_1 depends on $W(1)$, via A, but not $W(2)$, and g_2 depends on both $W(1)$ and $W(2)$. When $k = 3$, g_2 depends on $W_3 = [W(1),W(2),W(3)]$ and g_1 only on $W(1)$ and $W(2)$. In general if $k \geq 2$, g_2 depends on W_k

but g_1 depends only on W_{k-1} and not W_k, this means the conditional density of $W(k)$ given W_{k-1} is g_2.

The above consideration give the following theorem.

<u>THEOREM 1</u>

If $\{Y(t): t = 0, \pm 1, \pm 2, \ldots\}$, is a AR(p) process with unknown parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_p)' \epsilon R^p$ and $\tau > 0$, S_n is a sample of n observations, W_k a sample of k future observations, $Y(0), Y(-1), \ldots, Y(1-p)$ are known real constants, and if the prior distribution of θ and τ is a normal-gamma density with known parameters $\mu \epsilon R^p$, P a positive definite symmetric matrix, $\alpha > 0$, and $\beta > 0$, then the predictive distribution of W_k is as follows:

(a) If $k = 1$, the predictive distribution of $W(1)$ is given by (16) with $k = 1$, which is a t density with $n+2\alpha$ degrees of freedom, location

$$E[W(1)/s] = D_1^{-1} E_1 \tag{24}$$

and precision

$$P[W(1)/S_n] = \frac{(n+2\alpha)D_1}{F_1 - E_1' D_1^{-1} E_1} \tag{25}$$

where

$$D_1 = 1 - G_o' A^{-1} G_o ,$$

$$E_1 = G_o' A^{-1} (B_1 + P\mu),$$

$$F_1 = C - W^2(1) - (B_1 + P\mu)' A^{-1} (B_1 + P\mu) ,$$

and

$$G_o = [W(0), W(-1), \ldots, W(1-P)]' .$$

(b) If $K = 2$, the predictive distribution of W_2 is such that the margin. distr' .ion of $W(1)$ is given by (a) above

and the conditional distribution of W(2) given W(1) is given by (16) with $k = 2$, which is a t density with $n+2\alpha$ degrees of freedom, location

$$E[W(2)/W(1),S_n] = D_2^{-1}E_2 . \tag{26}$$

and precision.

$$P[W(2)/W(1),S_n] = \frac{(n+2\alpha)D_2}{F_2-E_2D_2^{-1}E_2} , \tag{27}$$

where,

$$D_2 = 1 - G_{-1}A^{-1}G_{-1}$$

$$E_2 = G_{-1}\ A^{-1}[B_1+\mathbb{P}\mu + W(1)G_o] ,$$

$$F_2 = C - W^2(2) - [B_1+\mathbb{P}\mu+W(1)G_o]'A^{-1}[B_1+\mathbb{P}\mu+W(1)G_o] ,$$

and

$$G_1 = [W(-1),W(-2),\ldots,W(2-p)]' .$$

(c) If $K = 3$, the predictive distribution of W_3 is such that the marginal distribution of W(1) is given by (a), the conditional distribution of W(2) given W(1) is given above by (b), and the conditional density of W(3) given W(1) and W(2) is the t-density (16) with $K = 3$, with $n+2\alpha$ degrees of freedom location

$$E[W(3)/W(1),W(2),S_n] = D_3^{-1}E_3 , \tag{28}$$

and precision

$$P[W(3)/W(1),W(2),S_n] = \frac{(n+2\alpha)D_3}{F_3-E_3D_3^{-1}E_3} \tag{29}$$

where

$$D_3 = 1 - G'_{-2}A^{-1}G_{-2} ,$$

$$E_3 = G'_{-2}A^{-1}[B_1+\mathbb{P}\mu+W(1)G_o+W(2)G_{-1}] ,$$

$$F_3 = C - W^2(3) - [B_1+\mathbb{P}\mu+W(1)G_o+W(2)G_{-1}]'A^{-1}[B_1+\mathbb{P}\mu+W(1)G_o+W(2)G_{-1}],$$

and

$$\underline{G}_2 = [W(-2),\ldots,W(3-p)]'.$$

(d) In general if $K \geq 2$, the conditional predictive distribution of $W(k)$ given $W_{k-1} = [W(1),W(2),\ldots,W(k-1)]'$ is t with $n+2\alpha$ degrees of freedom, location.

$$E[W(k)/W_{k-1},S_n] = D_k^{-1} E_k \tag{30}$$

and precision

$$P[W(k)/W_{k-1},S_n] = \frac{(n+2\alpha)D_k}{F_k - E_k D_k^{-1} E_k} \tag{31}$$

where

$$D_k = 1 - G'_{-(k-1)} A^{-1} G_{-(k-1)} \ ,$$

$$E_k = G'_{-(k-1)} A^{-1} [B_1 + \mathbb{P}\mu + \sum_{j=I}^{k-1} W(j) G_{-(j-1)}] \ ,$$

$$F_k = C - W^2(k) - [B_1 + \mathbb{P}\mu + \sum_{j=1}^{k-1} W(j) G_{-(j-1)}]' A^{-1} [B_1 + \mathbb{P}\mu + \sum_{j=1}^{k-1} W(j) G_{-(j-1)}],$$

where for $i = 0,1,2,\ldots$

$$G_{-i} = [W(-i),W(-i-1),\ldots W(i+1-p)]' \ .$$

Thus it is seen that the predictive density of W_k maybe expressed as the product of k univariate t densities, namely the marginal density of $W(1)$, given by (a). the conditional predictive density of $W(2)$ given $W(1)$, given above by (b), and so on; however, the predictive density of W_k is not the standard k-variate multivariate t density as defined by DeGroot (1970), Press (1972), and Zellner (1971).

What is the predictive distribution of W_k if one uses Jeffreys' prior density ξ_1, (2), in lieu of the conjugate prior density ξ_2, (3),? Fortunately, one may revise the previous theorem and thereby produce the predictive distribution.

<u>**THEOREM 2**</u>

If$\{Y(t): t = 0,\pm 1,\pm 2,\ldots\}$ is an AR(p) process with unknown

parameters $\theta \epsilon R^p$ and $\tau > 0$, S_n a sample of n observations, $n > p$, $y(0), Y(-1), \ldots, Y(1-p)$ known real constants, and if the prior density of θ and τ is $\xi(1),(2)$, the predictive distribution of W_k is given by Theorem 1 by letting $\mathbb{P} \rightarrow 0(p \times p), \alpha \rightarrow -p/2$, and $\beta \rightarrow 0$ in equations (24) through (31).

In particular, consider a first order model, $p = 1$, with Jeffreys' prior density ξ_1 and a one-step ahead forecast, $k = 1$, then what is the predictive density of $W(1)$? According to Theorem 2, part (a) of Theorem 1 gives the solution with $\alpha \rightarrow -\frac{1}{2}$, $\beta \rightarrow 0$, and $\mathbb{P} \rightarrow 0$ substituted into equations (24) and (25). This gives a t distribution for $W(1)$ with n-1 degrees of freedom, location

$$E[W(1)/S_n] = \frac{Y(n) \sum_0^n Y(t)Y(t-1)}{\sum_0^{n-1} y^2(t)}, \tag{32}$$

and precision

$$P[W(1)/S_n] = \frac{(n-1) \sum_0^{n-1} Y^2(t)}{[\sum_1^n Y^2(t)][\sum_0^n Y^2(t)] - [\sum_1^n Y(t)Y(t-1)]^2 [\sum_0^n Y^2(t)][\sum_0^{n-1} Y^2(t)]^{-1}} \tag{33}$$

Using the vague prior, the mean of the predictive distribution of $W(1)$ is

$$\hat{Y}(n+1) = \hat{\theta} Y(n),$$

where the posterior mean of θ

$$\hat{\theta} = \sum_0^n Y(t)Y(t-1) / \sum_0^{n-1} Y^2(t), \tag{34}$$

is an estimate of θ, the autoregressive parameter, which when $|\theta| \leq 1$, is the autocorrelation between successive observations.

A straightforward way to predict $Y(n+1)$ is to note

$Y(n+1) = \theta Y(n) + \varepsilon(n+1)$,

thus $EY(n+1) = Y(n)$, where the average E is taken with respect to $\varepsilon(n+1)$ given $Y(n)$, then $EY(n+1)$ is estimated by $\theta^* Y(n)$, where θ^* is some estimate of θ, say the mean , (34), of the posterior distribution of θ. Hence the Bayesian way of point forecasting with the predictive mean conforms to the usual way one would attempt to solve the problem.

If one would want to forecast $W(1)$, one could use an interval prediction based on the predictive distribution of $W(1)$, which is t with n-1 degrees of freedom, location given by (32) and precision given by (33). The predictive variance of $W(1)$ is

$$\text{Var}\,[W(1)/S_n] = (n-1)(n-3)^{-1}P^{-1}[W(1)/S_n] \;, \tag{34}$$

thus

$$E[W(1)/S_n] \pm t_{\alpha/2,n-1}\sqrt{\text{var}[W(1)/S_n]} \tag{35}$$

is a $1-\gamma$, $0 < \gamma < 1$, prediction interval of $Y(n+1)$ and is easily computed with the aid of student's t tables. The intervals have the HPD (highest posterior density) property explained by Box and Tiao (1965).

Land (1981) gives some examples of one and two-step ahead forecasting, via Theorem 2, with an AR(1) process and Jeffreys' prior distribution.

4. A Numerical Study

In this section of the study, an AR(1) model

$$Y(t) = \phi Y(t-1) + \varepsilon(t) \;,\; t = 1,2,\ldots,n \tag{36}$$

is considered, where $Y(t)$ is the observation at time t, $Y(0)$ is a

known constant, $\phi \epsilon R$ is the unknown autoregressive parameter and the $\varepsilon(t)$, $t = 1,2,\ldots,n$ are n.i.d. $(0,\tau)$, where $\tau > 0$ is unknown. Suppose the prior density for ϕ and τ is

$$\xi(\phi,\tau) \propto \tau^{\frac{1}{2}} e^{-\frac{\tau}{2}(\phi-\mu)^2 \mathbb{P}} \tau^{\alpha-1} e^{-\tau\beta}, \phi \epsilon R, \tau > 0 , \tag{37}$$

which is a normal-gamma density with parameters $\mu \epsilon R$, $\mathbb{P} > 0$, α and $\beta > 0$. The marginal prior density of τ is gamma with parameters $\alpha > 0$ and $\beta > 0$ and the marginal prior density of ϕ is

$$\xi(\phi) \propto [2\beta + (\theta-\mu)^2 \mathbb{P}]^{-(1+2\alpha)/2}, \phi \epsilon R \tag{38}$$

thus apriori, ϕ has a t distribution with 2α degrees of freedom, location μ, precision $(2\alpha)\mathbb{P}(2\beta)^{-1}$ and variance $\beta(\alpha-1)^{-1}\mathbb{P}^{-1}$, when $\alpha > 1$.

Suppose one believes that the process is 'almost' stationary, then one would want 'most' of the marginal prior probability distribution of ϕ to be concentrated over $(-1,+1)$. For example, suppose one wants $\phi \epsilon (-1,1)$ with a preassigned probability of $1-\gamma$, where $0 < \gamma < 1$, then, assuming α, β, and μ are fixed, one would choose $\mathbb{P}$ such that

$$\mathbb{P} = \beta t^2_{\gamma/2,2\alpha}(1-\mu)^{-2}(\alpha-1)^{-1}, |\mu| < 1, \alpha > 1, \tag{39}$$

where $t_{\gamma/2,2\alpha}$ is the upper $100(\gamma/2)\%$ point of the t-distribution with 2α degrees of freedom. Hence α and β are chosen to express the prior opinion of τ, μ is chosen as one initial guess of the value of ϕ, and then $\mathbb{P}$ is determined from (39).

In this way, one may express one's prior opinion of an almost (γ close to 1) stationary AR(1) process (36).

Now, suppose we want to forecast a future observation $Y(n+1)$

based on a sample S_n, when the observations were generated from an AR(1) process, which is almost stationary. Clearly, part (A) of theorem 1 applies and the following tables were computed using formulas (24) and (25) of that theorem.

Using the normal-gamma prior (37) for ϕ and τ, samples S_n, where n(= 25,50,100,750), were generated with ϕ = 0.0,.50,.75, and .90, τ = 1, and Y(0) = 0. The parameters of the prior distribution were α = 10, β = α - 1, μ = 0.0,0.25,0.5,0.75, and 0.90, and P was determined from (39) with γ = .05, that is, 95% of the marginal prior distribution of ϕ is concentrated over (-1,1), with a prior mean of μ and a prior variance of P^{-1}.

Consider Table 2, where sample S_n, n = 25,50,100,750 was generated from the AR(1) model Y(t) = .25Y(t-1) + ε(t), with Y(0) = 0 and ε(t) ~ n(0,1). The parameters of the prior distribution were chosen as α = 10, β= 9, μ = 0,.25,.50,.75,.9 , which determined P from (39) with γ = .05 . These tables consist of three parts, namely the prior, posterior, and predictive information, thus the first row corresponds to n = 25, μ = 0, P^{-1} = .2298, $E(\tau) = \alpha/\beta$ = 1.1111 , $var(\tau) = \alpha/\beta^2$ = .1235, the marginal posterior mean of ϕ is .3004, the marginal variance of ϕ is .0299, the mean of the predictive distribution of Y(n+1) is .4189, which is calculated from (24) and the predictive variance of Y(n+1) is 1.2954, which was computed from (25). Each table corresponds to a particular AR(1) model and are as follows.

TABLE 1

$AR(1), Y_t = \varepsilon_t, \varepsilon_t \sim i.i.d\ N(0,1), y_o = 0$, N-G prior $\alpha = 10, \beta = 9$

ϕ	n	Prior Information				Posterior Information				Predictive Information	
		$E(\phi)$	$V(\phi)$	$E(\tau)$	$V(\tau)$	$E(\phi/S)$	$V(\phi/S)$	$E(\tau/S)$	$V(\tau/S)$	$E(Y_{n+1})$	$V(Y_{n+1})$
0.00	25	0	.2298	1.1111	.1235	.1020	.0326	.8548	.0325	.1373	1.2834
	50	0	.2298	1.1111	.1235	.1805	.0180	.9211	.0242	-.2759	1.1596
	100	0	.2298	1.1111	.1235	.1425	.0094	.9968	.0165	.7584	1.0250
	750	0	.2298	1.1111	.1235	.0618	.0013	1.0216	.0027	.8453	.9868
0.25	25	.25	.1293	1.1111	.1235	.1408	.0299	.8538	.0324	.1885	1.2900
	50	.25	.1293	1.1111	.1235	.2008	.0171	.9227	.0243	-.3068	1.1554
	100	.25	.1293	1.1111	.1235	.1556	.0091	.9949	.0165	.8273	1.0247
	750	.25	.1293	1.1111	.1235	.0662	.0013	1.0212	.0027	.0896	.9872
0.50	25	.5	.0575	1.1111	.1235	.2476	.0269	.8290	.0305	.3334	1.3076
	50	.5	.0575	1.1111	.1235	.2651	.0151	.9080	.0236	-.4051	1.1690
	100	.5	.0575	1.1111	.1235	.1986	.0085	.9805	.0160	.1057	1.0396
	750	.5	.0575	1.1111	.1235	.0527	.0013	1.0168	.0027	.1065	.9914
0.75	25	.75	.0166	1.1111	.1235	.5450	.0138	.7280	.0237	.7339	1.4581
	50	.75	.0166	1.1111	.1235	.4982	.0099	.8179	.0191	-.7613	1.2817
	100	.75	.0166	1.1111	.1235	.3888	.0066	.8845	.0130	.2070	1.1516
	750	.75	.0166	1.1111	.1235	.1036	.0013	.9802	.0025	.2093	1.0281
0.90	25	.9	.0023	1.1111	.1235	.8443	.0036	.6285	.0176	1.1370	1.6215
	50	.9	.0023	1.1111	.1235	.8174	.0030	.6816	.0137	-1.2482	1.4955
	100	.9	.0023	1.1111	.1235	.7543	.0026	.7136	.0085	.4015	1.4258
	750	.9	.0023	1.1111	.1235	.3617	.0011	.8065	.0017	.7318	1.2474

TABLE 2

AR(1), $Y_t = .25Y_{t-1} + \varepsilon_t ; \varepsilon_t \sim$ i.i.d $N(0,1), y_o = 0$, N-G. prior $\alpha = 10$, $\beta = 9$

μ	n	Prior Information				Posterior Information				Predictive Information	
		$E(\phi)$	$V(\phi)$	$E(\tau)$	$V(\tau)$	$E(\phi/S)$	$V(\phi/S)$	$E(\tau/S)$	$V(\tau/S)$	$E(Y_{n+1})$	$V(Y_{n+1})$
0.00	25	0	.2298	1.1111	.1235	.3004	.0299	.8458	.0318	.4189	1.2954
	50	0	.2298	1.1111	.1235	.3892	.0156	.9104	.0237	-.5178	1.1583
	100	0	.2298	1.1111	.1235	.3686	.0083	.9887	.0161	.2012	1.0311
	750	0	.2298	1.1111	.1235	.2803	.0012	1.0205	.0027	.4964	.9863
0.25	25	.25	.1293	1.1111	.1235	.3209	.0275	.8521	.0323	.4475	1.2815
	50	.25	.1293	1.1111	.1235	.3973	.0148	.9166	.0240	-.5286	1.1493
	100	.25	.1293	1.1111	.1235	.3738	.0080	.9827	.0164	.2047	1.0360
	750	.25	.1293	1.1111	.1235	.2815	.0012	1.0210	.0027	.4986	.9859
0.50	25	.5	.0575	1.1111	.1235	.3883	.0227	.8472	.0319	.5414	1.2786
	50	.5	.0575	1.1111	.1235	.4315	.0131	.9175	.0242	-.5741	1.1451
	100	.5	.0575	1.1111	.1235	.3968	.0075	.9920	.0164	.2172	1.0274
	750	.5	.0575	1.1111	.1235	.2865	.0012	1.0199	.0027	.5074	.9869
0.75	25	.75	.0144	1.1111	.1235	.6063	.0124	.7908	.0278	.8453	1.3476
	50	.75	.0144	1.1111	.1235	.5839	.0086	.8743	.0218	-.7768	1.1926
	100	.75	.0144	1.1111	.1235	.5170	.0057	.9469	.0149	.2831	1.1057
	750	.75	.0144	1.1111	.1235	.3196	.0012	1.0023	.0026	.5658	1.0039
0.90	25	.9	.0023	1.1111	.1235	.8558	.0031	.7074	.0222	1.1932	1.4855
	50	.9	.0023	1.1111	.1235	.8343	.0026	.7770	.0173	-1.1089	1.3295
	100	.9	.0023	1.1111	.1235	.7880	.0023	.8219	.0113	.4314	1.2380
	750	.9	.0023	1.1111	.1235	.5002	.0009	.8936	.0021	.8851	1.1249

TABLE 3

AR(1), $y_t = .5y_{t-1} + \epsilon_t$, $\epsilon_t \sim$ i.i.d $N(0,1)$, $y_o = 0$, N-G prior $\alpha = 10$, $\beta = 9$

μ	n	Prior Information				Posterior Information				Predictive Information	
		E(ϕ)	V(ϕ)	E(τ)	V(τ)	E(ϕ/S)	V(ϕ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1)
0.00	25	0	.2298	1.1111	.1235	.4854	.0254	.8316	.0307	.7905	1.3259
	50	0	.2298	1.1111	.1235	.5846	.0121	.8926	.0228	-.8737	1.1783
	100	0	.2298	1.1111	.1235	.5896	.0062	.9788	.0160	.3236	1.0408
	750	0	.2298	1.1111	.1235	.5208	.0010	1.0180	.0027	.8017	.9862
0.25	25	.25	.1293	1.1111	.1235	.4900	.0235	.8600	.0314	.7985	1.3070
	50	.25	.1293	1.1111	.1235	.5842	.0116	.8998	.0231	-.8399	1.1681
	100	.25	.1293	1.1111	.1235	.5893	.0061	.9838	.0161	.3232	1.0355
	750	.25	.1293	1.1111	.1235	.5209	.0010	1.0198	.0027	.8019	.9856
0.50	25	.5	.0575	1.1111	.1235	.5232	.0192	.8490	.0320	.8522	1.2869
	50	.5	.0575	1.1111	.1235	.5945	.0105	.9085	.0236	-.8547	1.1547
	100	.5	.0575	1.1111	.1235	.5952	.0057	.9900	.0163	.3265	1.0290
	750	.5	.0575	1.1111	.1235	.5226	.0010	1.0206	.0027	.8045	.9846
0.75	25	.75	.0166	1.1111	.1235	.6644	.0110	.8280	.0305	1.0821	1.2912
	50	.75	.0166	1.1111	.1235	.6722	.0071	.9020	.0232	-.9664	1.1560
	100	.75	.0166	1.1111	.1235	.6689	.0065	.9831	.0161	.3559	1.0357
	750	.75	.0166	1.1111	.1235	.5376	.0008	1.0161	.0027	.8277	.9889
0.90	25	.9	.0023	1.1111	.1235	.8656	.0028	.7690	.0263	1.6096	1.3684
	50	.9	.0023	1.1111	.1235	.8503	.0023	.8426	.0203	-1.2225	1.2265
	100	.9	.0023	1.1111	.1235	.8200	.0019	.9140	.0139	.4498	1.1132
	750	.9	.0023	1.1111	.1235	.6365	.0007	.9653	.0024	.9798	1.0403

TABLE 4

$AR(1), y_t = .75y_{t-1} + \varepsilon_t, \varepsilon_t \sim i.i.d\ N(0,1), y_o = 0$, N-G prior $\alpha=10$, $\beta=9$

M	n	Prior Information				Posterior Information				Predictive Information	
		E(φ)	V(φ)	E(τ)	V(τ)	E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1)
0.00	25	0	.2298	1.1111	.1235	.7231	.0168	.8119	.0293	1.6301	1.3742
	50	0	.2298	1.1111	.1235	.7987	.0069	.8743	.0218	-1.7879	1-2121
	100	0	.2298	1.1111	.1235	.8100	.0033	.9667	.0156	.2774	1.0526
	750	0	.2298	1.1111	.1235	.7591	.0006	1.0168	.0027	1.0167	.9871
0.25	25	.25	.1293	1.1111	.1235	.7167	.0159	.8200	.0298	1.6157	1.3520
	50	.25	.1293	1.1111	.1235	.7943	.0068	.8793	.0221	-1.7779	1.2046
	100	.25	.1293	1.1111	.1235	.8075	.0032	.9700	.0157	.2765	1.0482
	750	.25	.1293	1.1111	.1235	.7587	.0006	1.0175	.0027	1.0162	.9864
0.50	25	.5	.0575	1.1111	.1235	.7149	.0140	.8332	.0308	1.6116	1.3270
	50	.5	.0575	1.1111	.1235	.7893	.0063	.8881	.0225	-1.7668	1.1909
	100	.5	.0575	1.1111	.1235	.8046	.0031	.9760	.0159	.2755	1.0423
	750	.5	.0575	1.1111	.1235	.7584	.0006	1.0186	.0027	1.0158	.9853
0.75	25	.75	.0144	1.1111	.1235	.7584	.0087	.8487	.0320	1.7087	1.2722
	50	.75	.0144	1.1111	.1235	.7981	.0048	.9038	.0233	-1.7887	1.1632
	100	.75	.0144	1.1111	.1235	.8083	.0027	.9874	.0162	.2768	1.0302
	750	.75	.0144	1.1111	.1235	.7605	.0005	1.0201	.0027	1.0187	.9838
0.90	25	.9	.0023	1.1111	.1235	.8809	.0025	.8316	.0307	1.9858	1.2713
	50	.9	.0023	1.1111	.1235	.8779	.0019	.8926	.0230	-1.9651	1.1566
	100	.9	.0023	1.1111	.1235	.8668	.0014	.9805	.0160	.2969	1.0373
	750	.9	.0023	1.1111	.1235	.7887	.0005	1.011	.0027	1.0504	.9924

TABLE 5

AR(1), $y_t = .9y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim$ i.i.d N(0,1), $y_o = 0$, N-G prior $\alpha = 10$, $\beta = 9$

μ	n	Prior Information				Posterior Information				Predictive Information	
		E(φ)	V(φ)	E(τ)	V(τ)	E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1
0.00	25	0	.2298	1.1111	.1235	.9146	.0082	.7961	.0282	3.7187	1.4498
	50	0	.2298	1.1111	.1235	.9332	.0025	.8641	.0213	-2.0812	1.2035
	100	0	.2298	1.1111	.1235	.9282	.0013	.9549	.0152	.2090	1.0651
	750	0	.2298	1.1111	.1235	.9036	.0002	1.0152	.0027	.9146	.9878
0.25	25	.25	.1293	1.1111	.1235	.9073	.0080	.7996	.0286	3.6900	1.4611
	50	.25	.1293	1.1111	.1235	.9307	.0026	.8661	.0216	-2.0755	1.2007
	100	.25	.1293	1.1111	.1235	.9277	.0013	.9563	.0153	.2087	1.0615
	750	.25	.1293	1.1111	.1235	.9033	.0002	1.0155	.0027	.9142	.9875
0.50	25	.5	.0575	1.1111	.1235	.8958	.0075	.8062	.0289	3.0631	1.4216
	50	.5	.0575	1.1111	.1235	.9262	.0024	.8700	.0216	-2.0655	1.1951
	100	.5	.0575	1.1111	.1235	.9231	.0013	.9592	.0153	.2081	1.0603
	750	.5	.0575	1.1111	.1235	.9028	.0002	1.0163	.0027	.9137	.9869
0.75	25	.75	.0144	1.1111	.1235	.8814	.0056	.8253	.0303	3.5844	1.3508
	50	.75	.0144	1.1111	.1235	.9174	.0021	.8814	.0222	-2.0458	1.1785
	100	.75	.0144	1.1111	.1235	.9197	.0012	.9672	.0156	.2060	1.0515
	750	.75	.0144	1.1111	.1235	.9019	.0002	1.0178	.0027	.9128	.9853
0.90	25	.9	.0023	1.1111	.1235	.9105	.0021	.8495	.0321	3.7030	1.2663
	50	.9	.0023	1.1111	.1235	.9218	.0012	.9023	.0233	-2.0558	1.1471
	100	.9	.0023	1.1111	.1235	.9223	.0008	.9818	.0161	.2075	1.0358
	750	.9	.0023	1.1111	.1235	.9041	.0002	1.0199	.0027	.9150	.9832

What can be concluded from this numerical study? First, one sees the influence of the sample size n and the prior mean μ on the posterior mean and variance and the predictive mean and variance of $Y(n+1)$. They show the anticipated reduction in the posterior variance of ϕ and the predictive variance of the future observation as the sample size increases. As the prior mean of ϕ increases toward one, for the same AR(1) series (same value of ϕ), and the same value of n, series length, the posterior variance of ϕ decreases, because as μ increases to one, IP, the prior variance of ϕ decreases, as can be verified from (39). Of course, this should happen since, as the prior variance of ϕ becomes smaller, so should the posterior variance of ϕ.

Also the tables show, that for the same value of μ and n, the posterior variance of ϕ decreases as the true value of ϕ increases from 0 to .90 . Of course, this is anticipated from the theory of time series, because the usual estimator of ϕ has a large-sample variance of $n^{-1}(1 - \phi^2)$, see Box and Jenkins (1970).

Results for the predictive density show that its mean is highly influenced by the value of the last observation $Y(n)$ and its variance by the sample size, the variance decreasing as the sample size increases.

Given a particular AR(1) model, say that given by Table 4, one would expect the predictive variance to be the smallest when $\mu = .75$ and $n = 750$ and this is indeed the case. The same holds for the other four tables. For example with Table 1, (when the 'true' value of $\phi = 0$), the predictive variance is smallest when

$\mu = 0$ and $n = 750$, which is the largest sample size.

Predictive intervals for one-step ahead forecasts are relatively easy to find. Suppose the AR(1) model with $\phi = .5$ (Table 3) is used to generate 25 observations and one's prior belief is based on $\mu = .75$, $\alpha = 10$, $\beta = 9$, and the P value given by (39), i.e., one is assuming the process is stationary with a prior probability of 95%. A 90% prediction interval for Y(26) is $1.0821 \pm 1.2917\, t_{.05,65}$, which is (-0.817,2.9812), and this follows from part a of Theorem 1.

5. Comments and Conclusions

It has been shown that if an AR(p) is an adequate model for a time series and if prior information is expressed as either a conjugate prior density or a Jeffreys' vague improper density, the predictive distribution of k future observations is characterized by the product of k univariate t densities. Theorem 1 and 2 give the particular details of the predictive distribution.

These theorems provide one with a way to make point and interval forecasts of future observations. For example, the mean of the predictive distribution gives a point forecast, and the mean together with the variance of the predictive distribution allows one to construct interval forecasts. The theorems also give a way to make one-step, two-step, and other poly-step predictions.

The numerical study illustrates the sensitivity of the posterior and predictive mean and variance to the sample size and prior mean of the autoregressive parameter of five AR(1) models.

The calculations reveal that the variance of the one-step ahead forecasts is a minimum for the largest sample size and when the prior mean of the autoregressive parameter coincides with the 'true' value of that parameter. The prior distribution of the parameters of the AR(1) models were chosen to express near stationarity of the process.

It would be interesting to develop the predictive distribution of vector autoregressive processes. One would expect the forecasting distribution to be characterized by the product of multivariate t densities.

6. Acknowledgement

The authors are grateful to Mr. M. Al-Mahmeed, who did the calculations for Tables 1 through 5. The first author was supported by research contract N000 14-82-K-0292 from the Office of Naval Research.

References

Aitchison, J. and I. R. Dunsmore, 1975, Statistical prediction analysis (Cambridge, London)

Box, G. E. P. and Gwilym M. Jenkins, 1970, Time series analysis, forecasting, and control (Holden-Day, San Francisco).

Box, G. E. P. and G. C. Tiao, 1965, Multiparameter problems from a Bayesian point of view, The annals of mathematical statistics 36, 1468-1482.

Chow, Gregory C., 1975, Multiperiod predictions from stochastic difference equations by Bayesian methods, chapter 8 of Studies in Bayesian econometrics and statistics, edited by S. E. Fienberg and A. Zellner, (North-Holland, Amsterdam).

DeGroot, M. H., 1970, Optimal statistical decisions (McGraw Hill, New York).

Granger, C. W. J. and Paul Newbold, 1977, Forecasting economic time series (Academic Press, New York).

Harrison, P. J. and C. F. Stevens, 1976, Bayesian forecasting (with discussion), Journal of the royal statistical society, series B 38, 205-247.

Land, Margaret, Foster, 1981, Bayesian forecasting for switching regression and autoregressive processes, Ph.D. dissertation (Oklahoma State University, Stillwater).

Press, James, S., 1972, Applied Multivariate analysis (Holt, Rinehart and Winston, New York).

Zellner, A., 1971, An introduction to Bayesian inference in econometrics (John Wiley, New York).